DOCUMENT RESUME

ED 423 289 TM 029 096

AUTHOR Bergstrom, Betty A.; Lunz, Mary E.
TITLE Rating Scale Analysis: Gauging the Impact of Positively and Negatively Worded Items.
PUB DATE 1998-04-00
NOTE 11p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Diego, CA, April 13-17, 1998).
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Adults; *Item Response Theory; *Rating Scales; Research Methodology; Test Format; *Test Items
IDENTIFIERS *Positive Wording (Tests)

ABSTRACT

This paper addresses questions of whether positively- and negatively-worded items measure the same construct and whether the rating scale categories "strongly agree" to "strongly disagree" are used in the same way for both types of items. Item response theory (IRT), specifically the Andrich Rating Scale Model (B. Wright and G. Masters, 1982) is used to analyze a survey on job satisfaction. The methodology presented provides a strategy for exploring the effect of including positively- and negatively-worded items on a Likert-type survey. The Job Satisfaction Survey (JSS) (P. Spector, 1983; 1992) is a 36-item instrument designed to measure job satisfaction, which purports to measure nine aspects of job satisfaction and overall job satisfaction. Seventeen items are positively worded, and 19 are negatively worded. The survey was administered to 706 respondents from diverse work settings. Data were analyzed with the Andrich Rating Scale model. Analysis indicated that the positively and negatively worded items appeared to be measuring the same construct. These results confirm that, for these data, positively and negatively worded items can be scaled together. The IRT analysis provided a method for determining the impact of mixing positively and negatively worded statements on the same scale. (Contains two tables, two figures, and seven references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 423 289

Rating Scale Analysis: Gauging the Impact of Positively and Negatively Worded Items

Betty A. Bergstrom, Ph.D.
Computer Adaptive Technologies, Inc

Mary E. Lunz, Ph.D.
American Society of Clinical Pathologists

Paper presented at the Annual Meeting
American Educational Research Association
San Diego, California
April 1998

TM029096

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Betty Bergstrom

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Rating Scale Analysis:
Gauging the Impact of Positively and Negatively Worded Items

Traditionally, one of the useful tools for understanding the diversity of human attitude and experience has been the Likert-type survey instrument. Survey instruments are used to collect, focus and organize human thought about a particular construct of interest. Using a Likert-type instrument with set responses such as "strongly agree", "agree", "disagree" and "strongly disagree" enables the researcher to collect and analyze data from large samples. Survey items are often written in both a positively and negatively worded direction to reduce agreement response tendencies (Spector, 1992). Negatively worded items are reverse scored to place them on the same continuum as the positively worded items.

Classical analysis of survey data attempts to scale subjects on the construct of interest (Crocker & Algina, 1986; Banerji, Smith & Dedrick, 1997). Survey respondents are often placed on a continuum based on a raw score average obtained by arbitrarily assigning a quantitative number to a qualitative response (strongly disagree=1; strongly agree=4). While differences in how respondents use the scale are acknowledged, there is no way to account for these differences. Differences in the use of the rating scale influence the interpretation of results.

Item Response Theory (IRT) can be used to analyze survey data. 1) A linear scale is constructed that allows for consistent quantitative comparisons between subjects. 2) The difficulty of each item and the difficulty of each rating category or "step" are estimated on the same scale as the estimation of respondent measures. This method of analysis provides more

detail on the intensity of respondents feelings. IRT also allows the comparison of the structure of different rating scales.

This paper addresses the questions of whether positively and negatively worded items measure the same construct and whether the rating scale categories "strongly agree" to "strongly disagree" are used in the same way for both types of items. IRT, specifically the Andrich Rating Scale Model (Wright & Masters, 1982) is used to analyze a survey on job satisfaction. The methodology presented provides a strategy for exploring the effect of including positively and negatively worded items on a Likert-type survey instrument.

Instrument and Data Collection

The Job Satisfaction Survey (JSS) is a 36 item instrument designed to measure employee job satisfaction (See Appendix A). It was developed by Paul Spector (1983; 1992) and purports to measure nine aspects of job satisfaction including pay, promotion, supervision, benefits, contingent rewards, operating procedures, co-workers, nature of work and communication. In addition to measuring the nine subscales, the scale was designed to yield a good measure of overall satisfaction. Seventeen of the items were positively worded so "strongly agree" would indicate high job satisfaction. Nineteen of the items were negatively worded so "strongly disagree" would indicate high job satisfaction. Positively and negatively worded items were included in all nine subscales.

The survey was administered in 1995 as part of a longitudinal study of professionals in an allied health occupation. Participants were randomly chosen from the total population of certified professionals. Sample size was 706 respondents from diverse work settings and geographic locations.

Data Analysis

Data were analyzed with the Andrich Rating Scale Model (Wright & Masters, 1982). This IRT model estimates the probability that a survey respondent will respond to a particular item with a particular response category as:

$$\log[P_{nij}/P_{ni(j-1)}] = B_n - D_i - F_j$$

where P_{nij} is the probability of respondent *n* scoring in category *j* of item *i*.; $P_{ni(j-1)}$ is the probability of respondent *n* scoring in category *j-1* of item *i*; B_n is the measure of respondent *n*; D_i is the difficulty of item *i* and F_j is the difficulty of step *j*.

The Andrich Rating Scale Model scores orders categories as steps. For an item on an attitude survey such as the JSS, completing the *j*th step can be thought of as choosing the *j*th step over the (*j*-1) step. Thus a respondent who chooses to Agree with a statement on the JSS when given the ordered categories

Strongly Disagree	*Disagree*	*Agree*	*Strongly Agree*
1	2	3	4

can be considered to have chosen *Disagree* over *Strongly Disagree* (first step taken) and also *Agree* over *Disagree* (second step taken), but to have failed to have chosen *Strongly Agree* over *Agree* (third step not taken) (Wright & Masters, 1982 p. 48).

Negatively worded items were re-coded (Strongly disagree (4), Disagree (3), Agree (2) and Strongly Disagree (1)). All 36 items were analyzed and estimates of respondent measures, item difficulty, standard errors of measure, fit statistics and reliability estimates were obtained (Bigsteps, 1997). This initial analysis indicated that the 36 items did represent a single unidimensional construct (job satisfaction) despite the categorization of the instrument into

additional subtests (See Bergstrom & Lunz, 1998). Then the data were re-analyzed, grouping the positively worded items on one rating scale and the negatively worded items on a separate scale. Lastly, estimates of respondent satisfaction on the two scales were compared.

Results

Tables 1 and 2 show a summary of the responses for the positively and the negatively worded items. The step measures (the calibrated difficulty of the step, relative to the prior step) indicate that respondents used the categories on the positively worded items from "strongly disagree" to "strongly agree" (-1.88, -.49, 2.37 logits) very similarly to the negatively worded items "strongly agree" to "strongly disagree" (-1.67, -.47, 2.15 logits).

The average measures (the average of the respondent measures that were modeled to produce the responses observed in the category) show a very similar pattern between the positive scale and the negative scale. Thurstone thresholds (the location of median probability) also show a similar pattern between the positive and negative scale.

TABLE 1 SUMMARY OF RESPONSES FOR THE POSITIVE ITEMS

CATEGORY OBSERVED LABEL		COUNT	AVGE MEASURE	INFIT MNSQ	OUTFIT MNSQ	STEP MEASURE	STEP ERROR	THURSTONE THRESHOLD
Strongly Disagree	1	1235	-1.28	1.06	1.09	NONE		
Disagree	2	3301	-.54	.84	.82	-1.88	.03	-2.07
Agree	3	5760	.67	.87	.89	-.49	.02	-.36
Strongly Agree	4	1589	1.43	1.17	1.13	2.37	.03	2.42

TABLE 2 SUMMARY OF RESPONSES FOR THE NEGATIVE ITEMS

CATEGORY OBSERVED LABEL		COUNT	AVGE MEASURE	INFIT MNSQ	OUTFIT MNSQ	STEP MEASURE	STEP ERROR	THURSTONE THRESHOLD
Strongly Agree	1	1708	-1.10	1.07	1.12	NONE		
Agree	2	4179	-.47	.91	.90	-1.67	.03	-1.89
Disagree	3	6110	.25	.93	.94	-.47	.02	-.32
Strongly Disagree	4	1300	1.01	1.06	1.07	2.15	.03	2.22

Figures 1 and 2 show the Response Category Probability Curves for the Positively and Negatively Worded Items.

CATEGORY PROBABILITIES: MODES - Step measures at intersections
PROBABILITY OF RESPONSE
RESPONDENT MEASURE [MINUS] ITEM DIFFICULTY

Figure 1. Response Category Probability Curves for positively worded items.

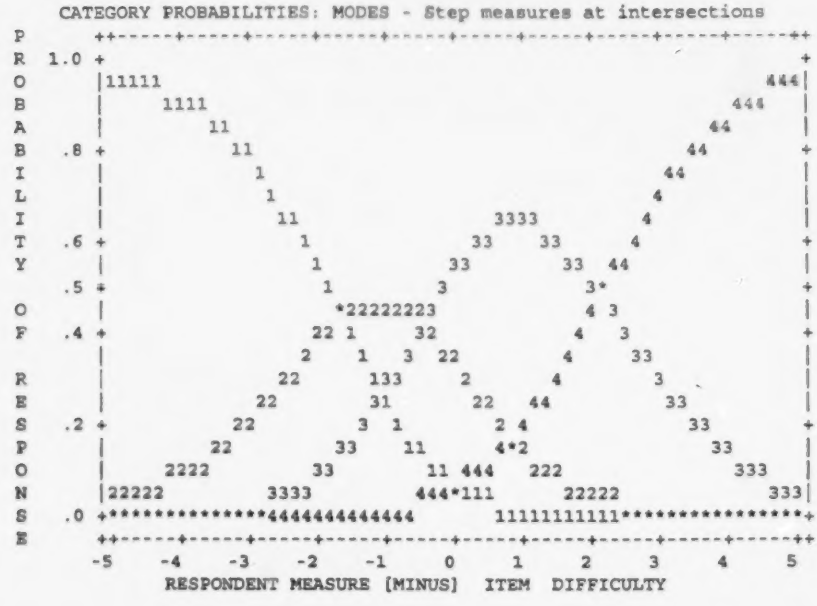
CATEGORY PROBABILITIES: MODES - Step measures at intersections
PROBABILITY OF RESPONSE
RESPONDENT MEASURE [MINUS] ITEM DIFFICULTY

Figure 2. Response Category Probability Curves for negatively worded items.

In Figures 1 and 2, for each scale, the probability of each response is shown across the measurement continuum. The probability that a particular respondent will respond with a particular category is determined by the difference between the estimated measure of the respondent and the calibrated difficulty of the item. Figures 1 and 2 show, for example, that the probability of a person responding with an "agree" statement on a positively worded item is very similar to the probability of responding to a "disagree" statement on a negatively worded item.

Additionally, respondent measures on the positively and negatively worded items were estimated separately. The correlation for the estimate of respondent measures for positively and negatively worded items was .77, and most respondents were measured comparably on the two types of items. These results indicate that comparable knowledge about the intensity of respondent job satisfaction was gained from positively and negatively worded items.

Reliability for the survey when all items were scaled together was .87. Reliability for the survey when positive and negative items were scaled separately also was .87. Item and respondent fit statistics did not improve when separate scales were used.

Conclusion

This analysis indicated that the two scales were used comparably and that the positively and negatively worded items appeared to be measuring the same construct. These results confirmed that for these data, positively and negatively worded items can be scaled together. The IRT analysis provided a method for determining the impact of mixing positively and negatively worded statements on the same scale.

The authors welcome comment on this paper:

Betty A. Bergstrom, Ph.D.
Computer Adaptive Technologies, Inc.
1007 Church Street
Evanston, IL 60201

bbergstrom@catinc.com

Mary E. Lunz
Measurement Resources, Inc.
Chicago, IL

Measresinc@aol.com

References

Banerji, M., Smith, R. M., & Dedrick, R. F. (1997). Dimensionality of an early childhood scale: Using Rasch analysis and confirmatory factor analysis. *Journal of Outcome Measurement*, *1*(1), 56-85.

Bergstrom, B & Lunz, M (April, 1998) *Measuring Job Satisfaction: Reliability of Subscale Analysis*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego, CA.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York, NY: CBS College Publishing.

Linacre, J. (1997). *BIGSTEPS* [Computer Program]. Chicago, IL: MESA Press.

Spector, P. E. (1983). Measurement of human service staff satisfaction: Development of the job satisfaction survey. *American Journal of Community Psychology*, *13*(6).

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Sage University Paper Series on Quantitative Applications in the Social Sciences, vol. 07-082. Newbury Park, CA: Sage Publications.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Appendix A

Job Satisfaction Survey

Indicate the strength of your agreement with each of the following points:

A—Strongly Disagree
B—Disagree
C—Agree
D—Strongly Disagree

1. I feel I am being paid a fair amount for the work I do.
2. There is really too little chance for promotion on my job.
3. My supervisor is quite competent in doing his/her job.
4. I am not satisfied with the benefits I receive.
5. When I do a good job, I receive the recognition for it that I should receive.
6. Many of our rules and procedures make doing a good job difficult.
7. I like the people I work with.
8. I sometimes feel my job is meaningless.
9. Communications seem good within this organization.
10. Raises are too few and far between.
11. Those who do well on the job stand a fair chance of being promoted.
12. My supervisor is unfair to me
13. The benefits we receive are as good as most other organizations offer.
14. I do not feel that the work I do is appreciated.
15. My efforts to do a good job are seldom blocked by red tape.
16. I find I have to work harder at my job than I should because of the incompetence of people I work with.
17. I like doing the things I do at work.
18. The goals of this organization are not clear to me.
19. I feel unappreciated by the organization when I think about what they pay me.
20. People get ahead as fast here as they do in other places.
21. My supervisor shows too little interest in the feelings of subordinates.
22. The benefit package we have is equitable.
23. There are too few rewards for those who work here.
24. I have too much to do at work.
25. I enjoy my co-workers.
26. I often feel that I do not know what is going on with the organization.
27. I feel a sense of pride in doing my job.
28. I feel satisfied with my chances for salary increases.
29. There are benefits we do not have which we should have.
30. I like my supervisor.
31. I have too much paperwork.
32. I don't feel my efforts are rewarded the way they should be.
33. I am satisfied with my chances for promotion.
34. There is too mcuh bickering and fighting at work.
35. My job is enjoyable.
36. Work assignments are often not fully explained.